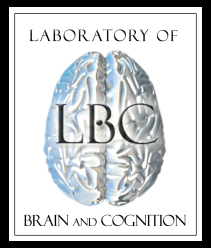
Decoding and Mind Reading with fMRI

07.19.11

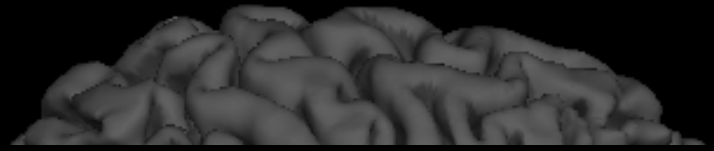
Chris Baker

Laboratory of Brain and Cognition, NIMH

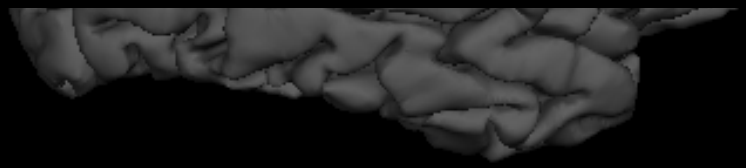


CBS – 60 Minutes, 2009

“mind reading” “thought identification”



What can we really do?



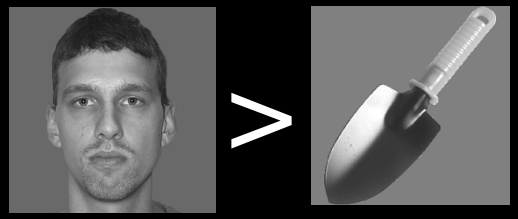
“prediction” “decoding”

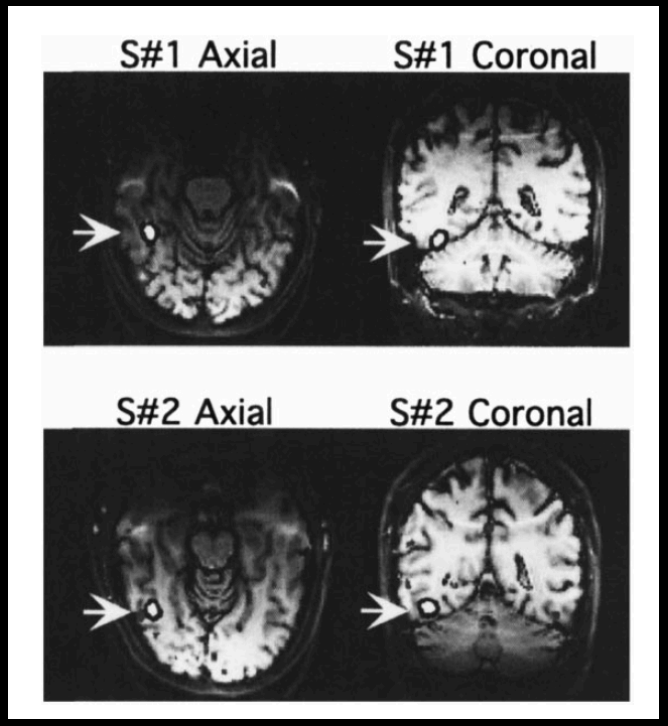
Let's Read Some Brains

1) Training
2) Test

I) Training

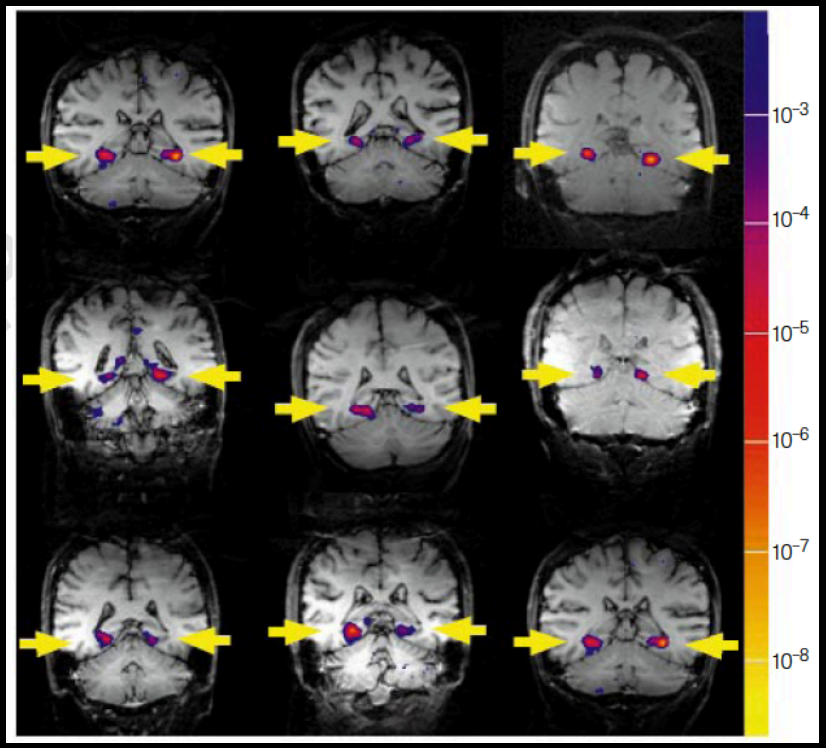
Face-selective cortex
(Fusiform Face Area, FFA)





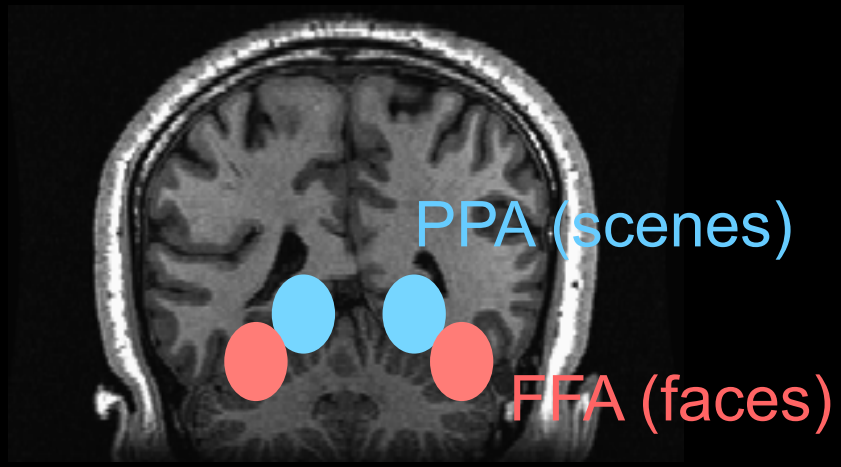
Kanwisher et al. (1997). *J. Neurosci.*, 17, 4302-4311

Scene-selective cortex
(Parahippocampal Place Area, PPA)



Epstein and Kanwisher (1998). *Nature*, 392, 598-601

I) Training



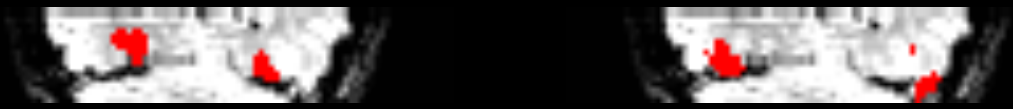
2) Test

S1

Faces

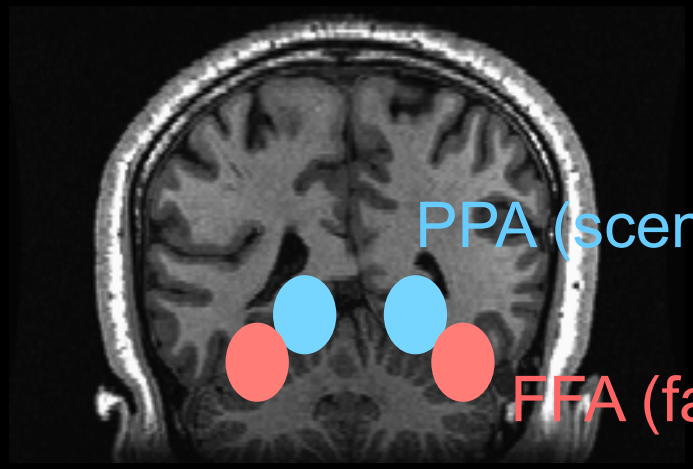
Scenes

S2



Scenes

Faces

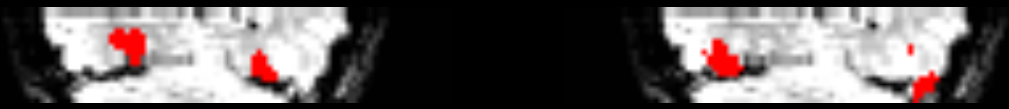


S1

Faces

Scenes

S2

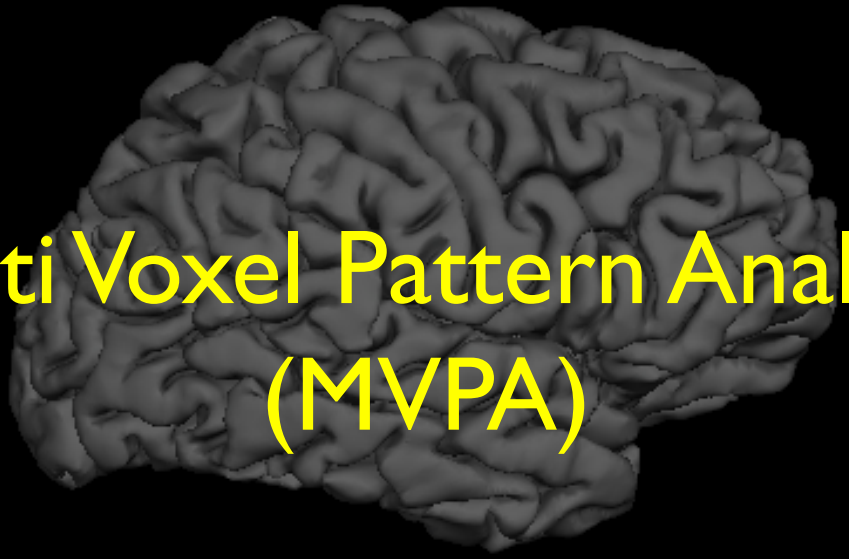


Scenes

Faces

"brain reading"

"classification"



Multi Voxel Pattern Analysis (MVPA)

"prediction"

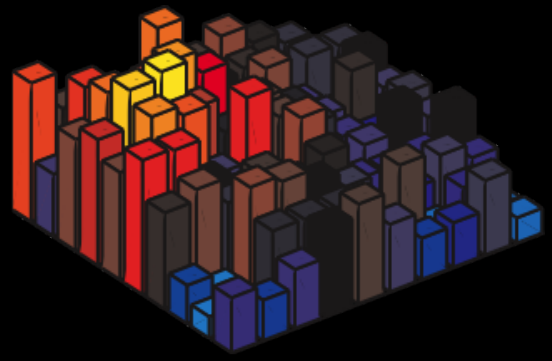
"decoding"

Univariate vs. Multivariate

- Classic fMRI analyses = univariate
  - Each voxel considered independently
- Multivariate
  - Responses of voxels considered jointly
  - Pattern of response

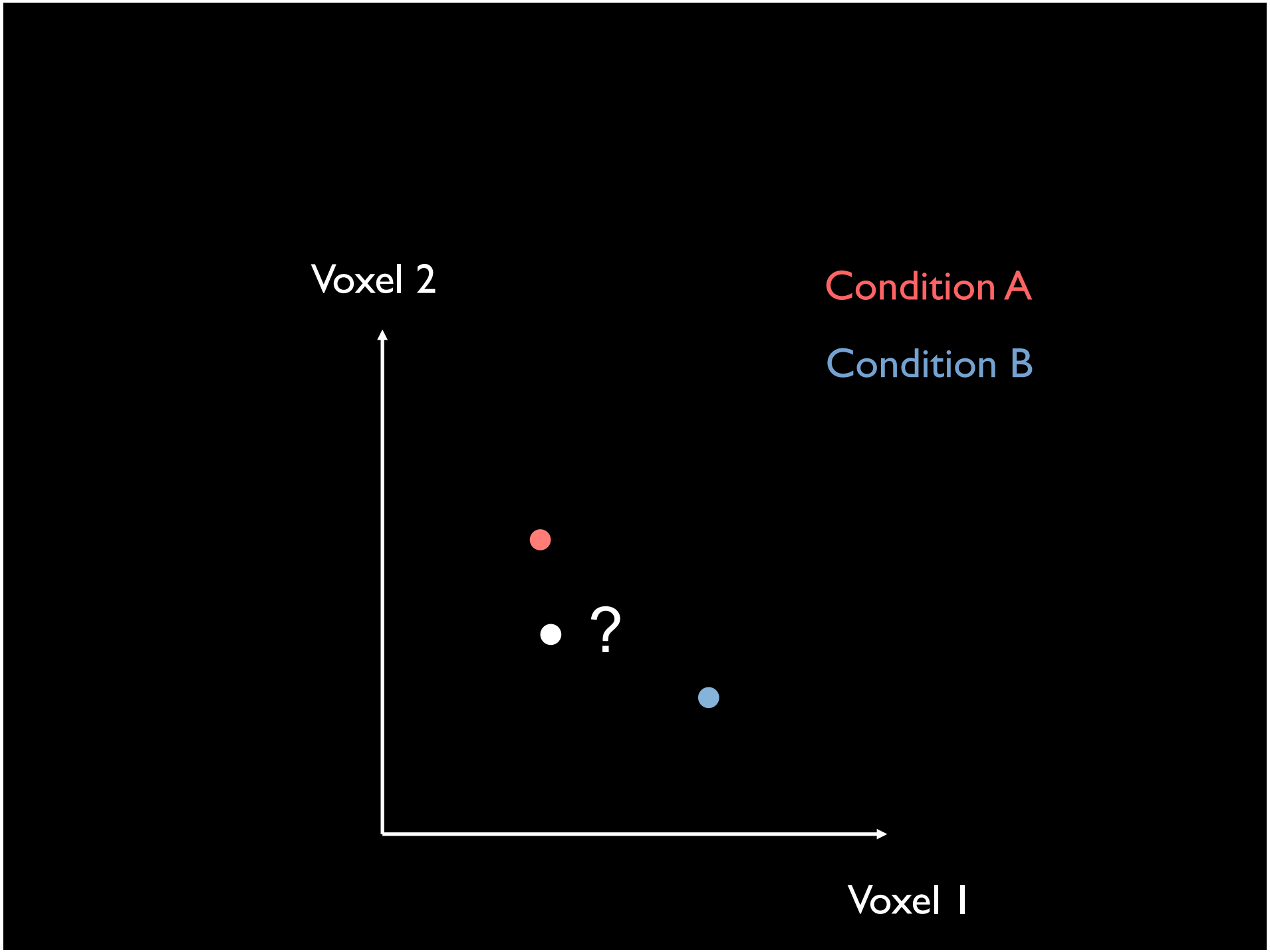
Condition 1

Condition 2

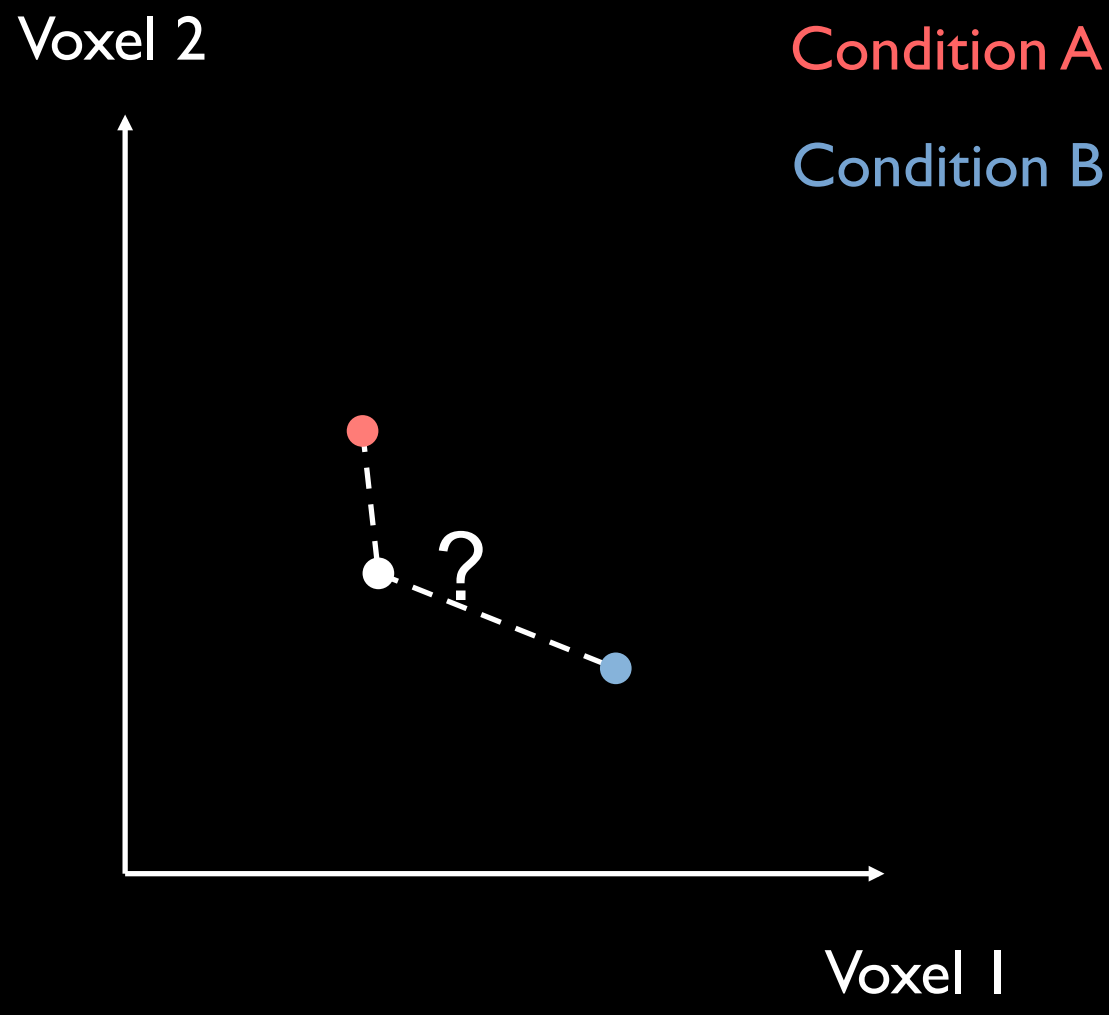


Condition 3

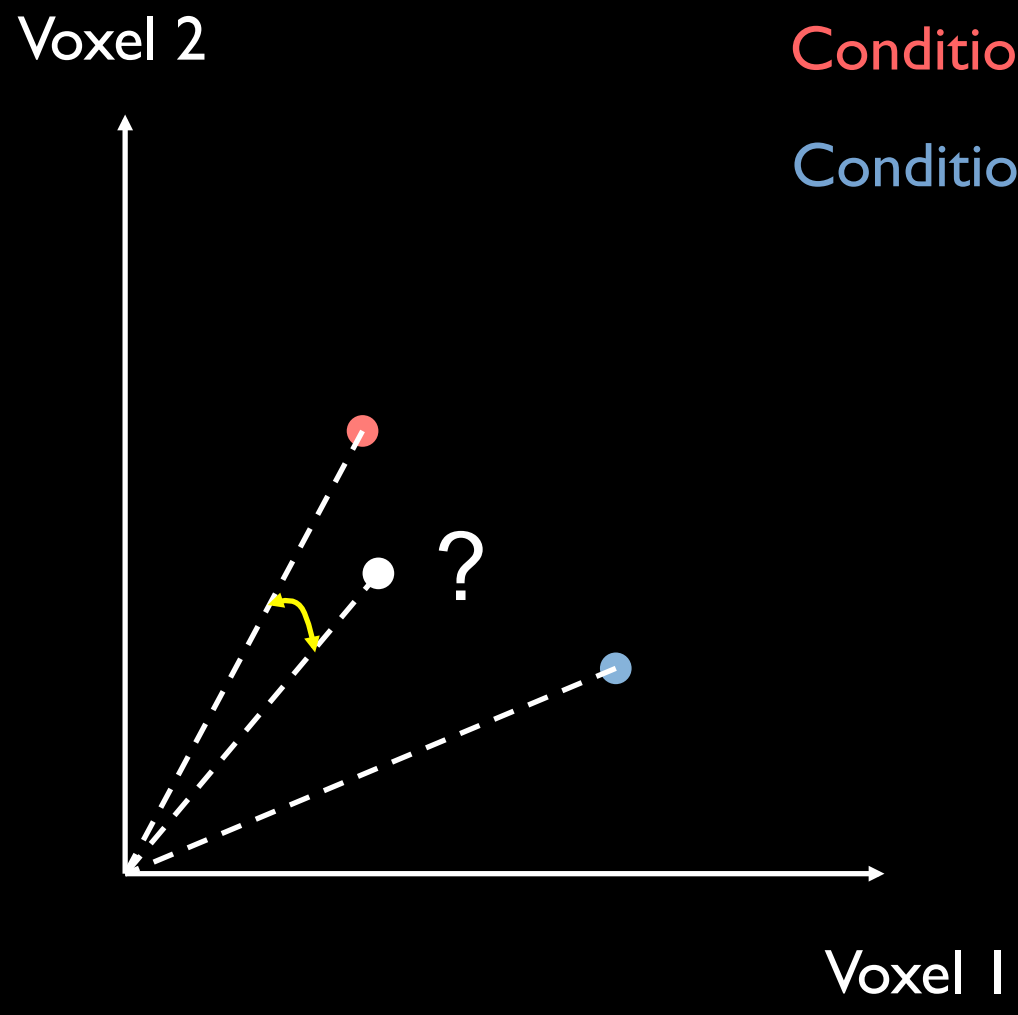
Condition n


Voxel 2
Condition A
Condition B
?
Voxel 1

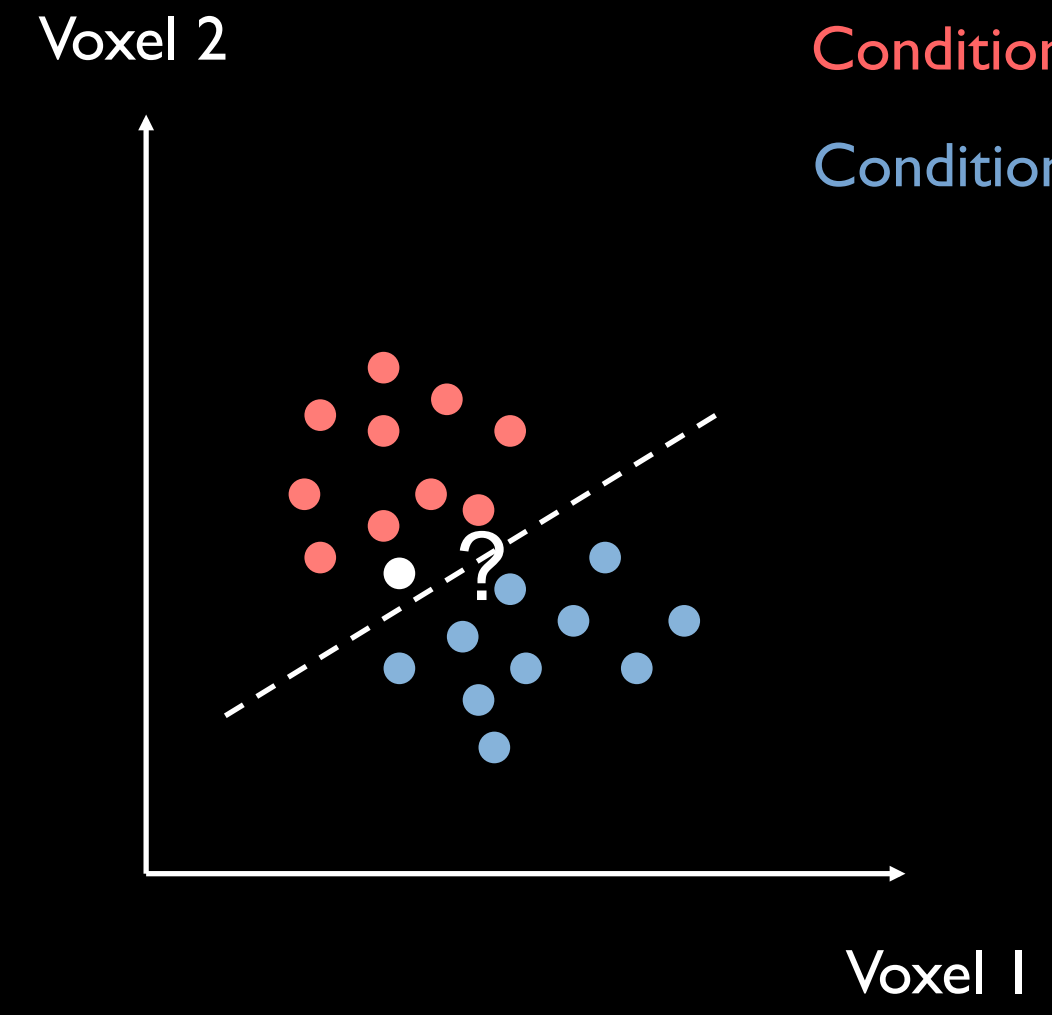
Euclidean Distance



Correlation



Support Vector Machine (SVM)



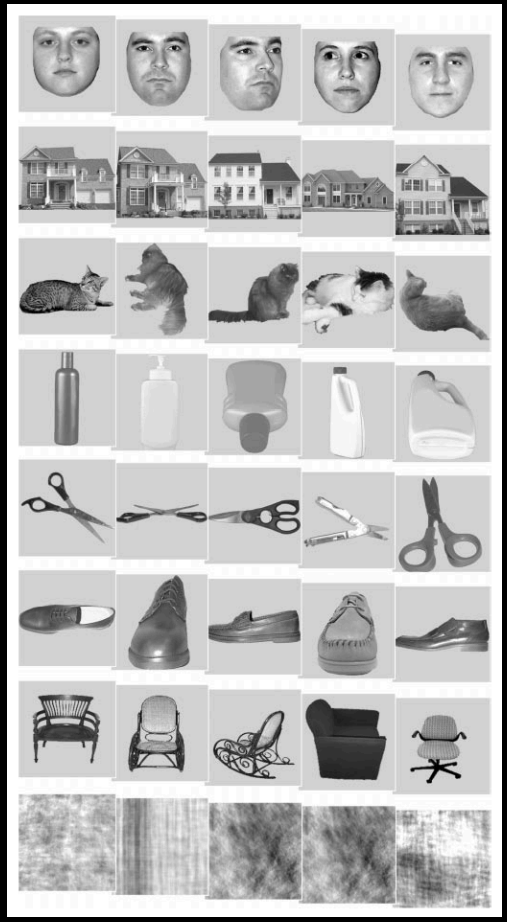
Linear Classifiers

- Euclidean distance
- Correlation
- Linear SVM
- Fisher Least Discriminant Analysis
- Neural networks (without hidden layer)
- Gaussian Naïve Bayes Classifiers

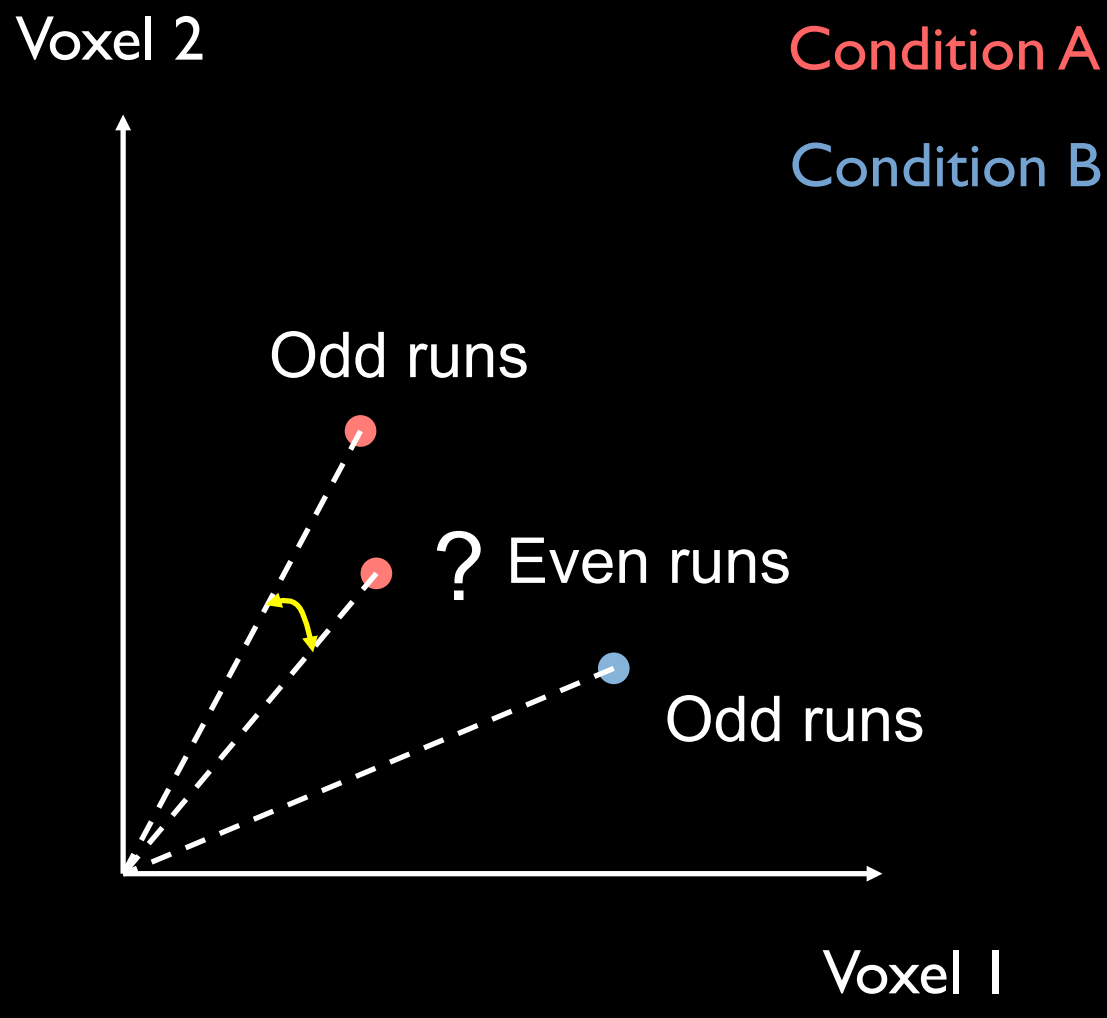
Non-linear classifiers increase risk of overfitting

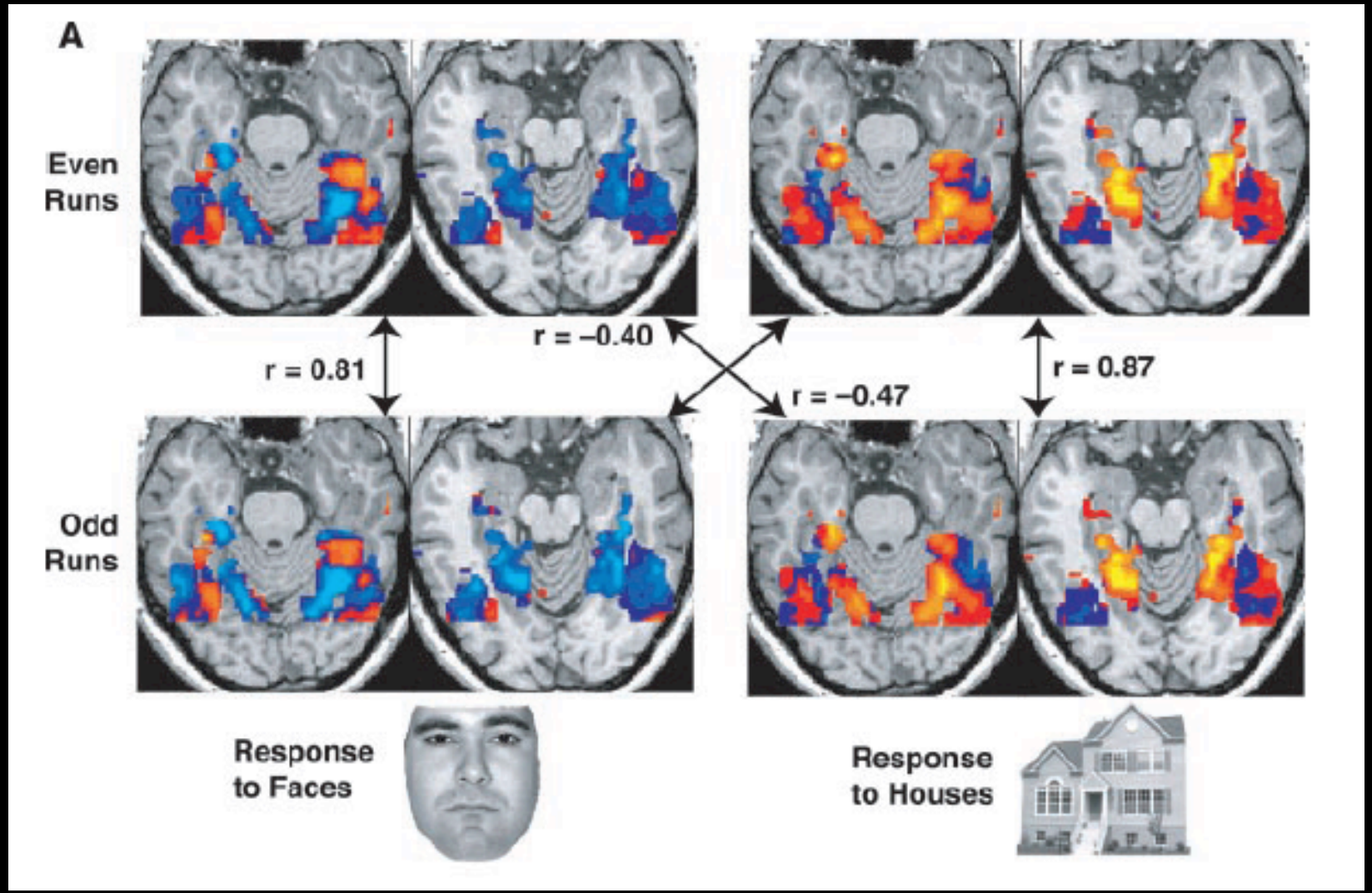
Object representations in ventral temporal cortex [Haxby et al. (2001)]

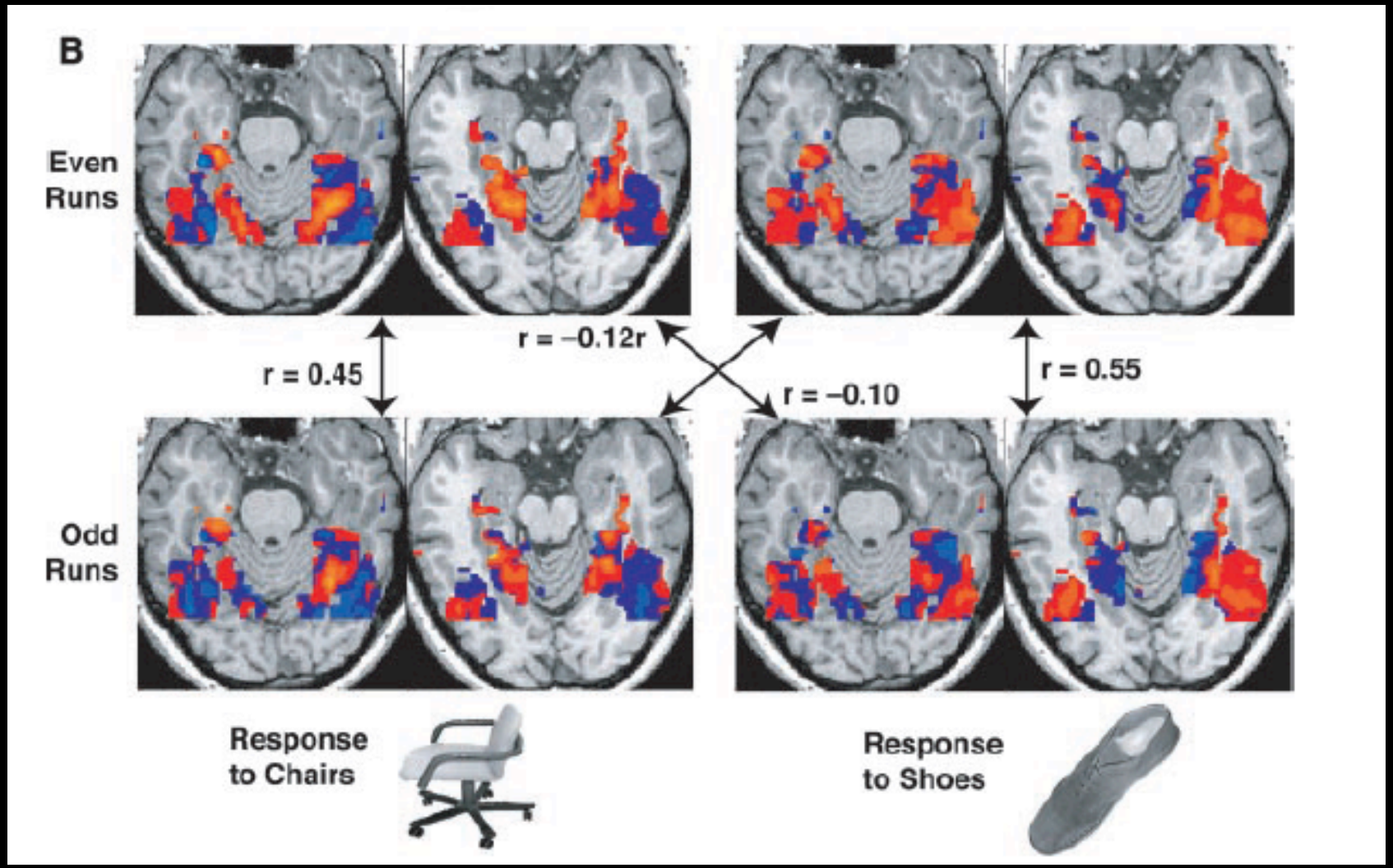
- Participants viewed blocks of images from 8 categories
- 1-back task
- Split-half correlation analysis



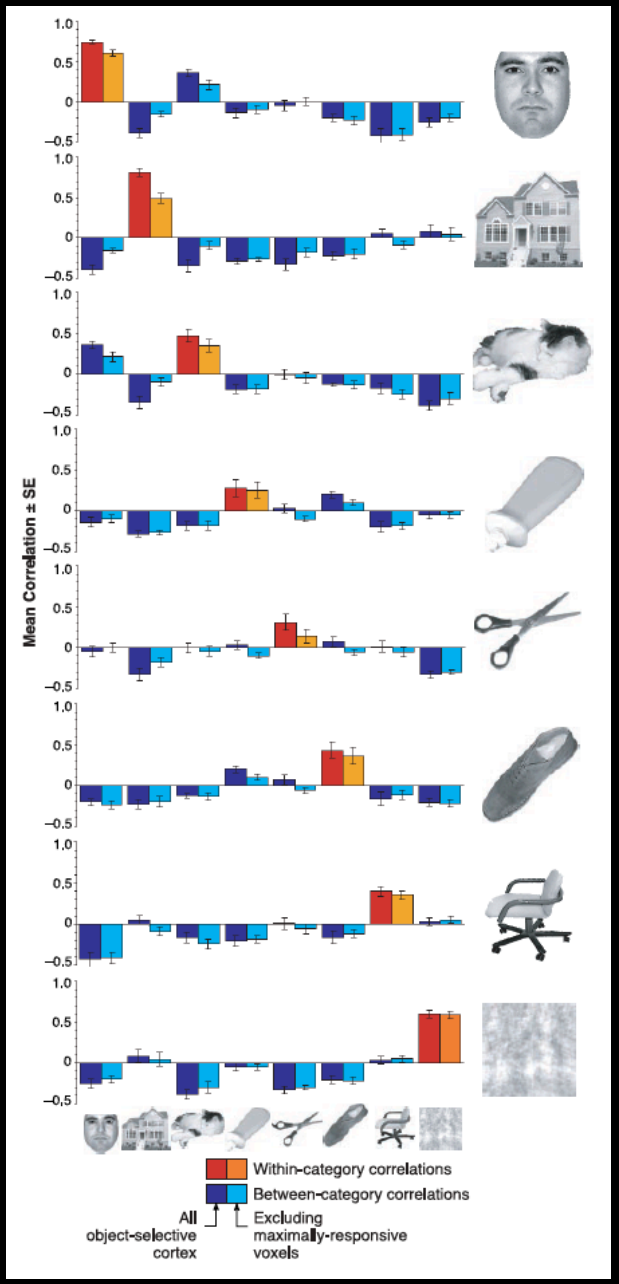
Correlation




A
Even Runs
Odd Runs
r = 0.81
r = −0.40
r = −0.47
r = 0.87
Response to Faces
Response to Houses


B
Even Runs
Odd Runs
r = 0.45
r = −0.12r
r = −0.10
r = 0.55
Response to Chairs
Response to Shoes

Higher within- than between-category correlations



Decoding Accuracy

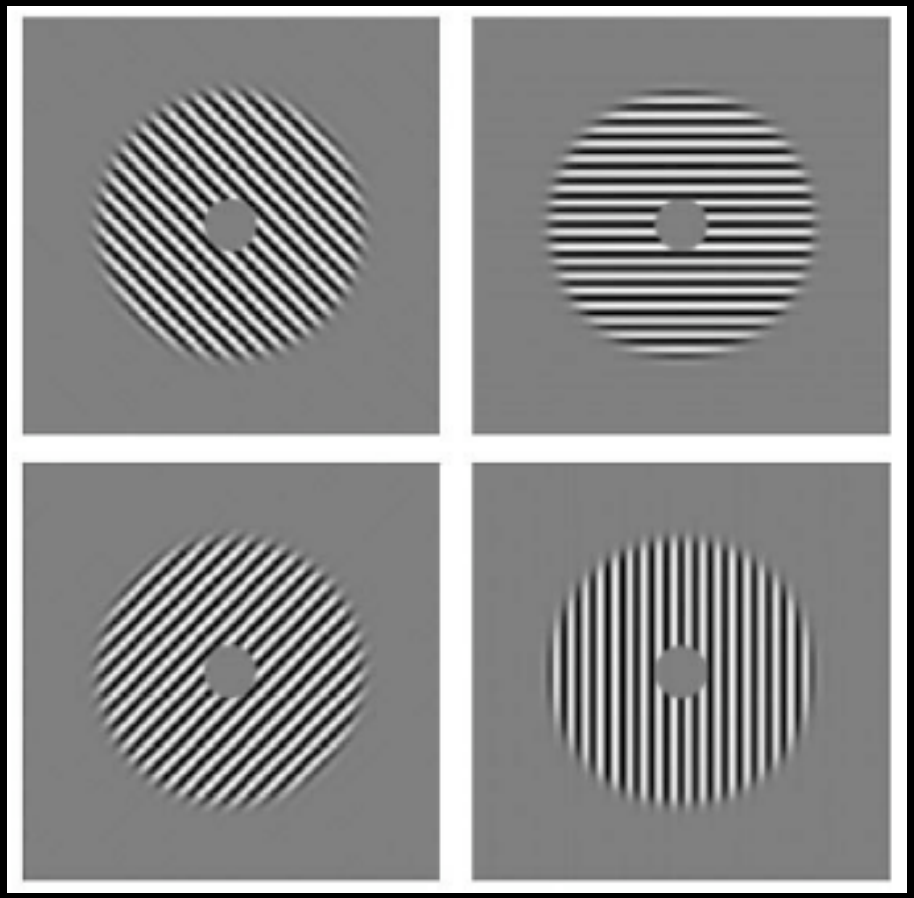
Table 1. Accuracy of identification of the category being viewed based on the patterns of response evoked in ventral temporal cortex. Accuracies are the percentage of comparisons between two categories that correctly identified which category was being viewed.

Region	Volume ($cm^3 \pm SE$)	Identification accuracy (%)							
		Faces	Houses	Cats	Bottles	Scissors	Shoes	Chairs	Scrambled
All ventral temporal object-selective cortex	22.9 ± 2.8	100***	100***	98 ± 2***	90 ± 6***	92 ± 6***	92 ± 7***	96 ± 2***	100***
Minus regions that were maximally responsive to categories being compared	15.4 ± 1.8	100***	100***	95 ± 2***	89 ± 6***	85 ± 9**	90 ± 8**	98 ± 1***	100***
Regions maximally responsive to:									
Faces	3.1 ± 0.9	94 ± 7***	99 ± 1***	76 ± 13*	81 ± 14*	77 ± 9*	70 ± 16	77 ± 11*	92 ± 7***
Houses	9.6 ± 1.8	100***	100***	88 ± 5***	85 ± 10**	81 ± 6**	96 ± 2***	94 ± 3***	100***
Cats	2.6 ± 0.4	96 ± 4***	96 ± 2***	82 ± 8**	65 ± 11	69 ± 5**	76 ± 9*	95 ± 4***	100***
Small objects	6.9 ± 1.1	100***	100***	95 ± 3***	83 ± 7**	92 ± 8**	94 ± 6***	90 ± 6***	96 ± 4***

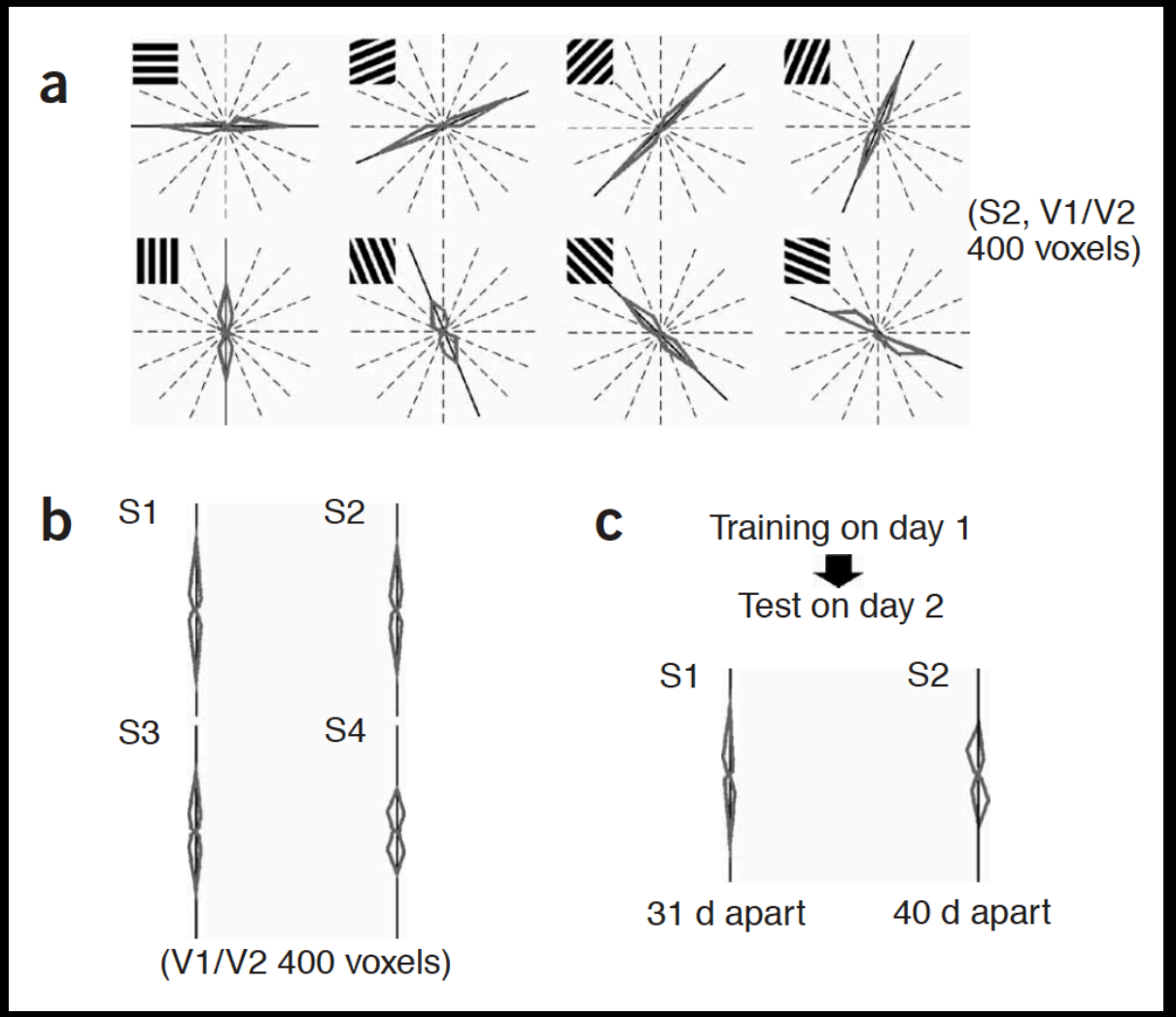
Differs from chance (50%): *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Decoding Orientation in Early Visual Cortex [Kamitani and Tong (2005)]

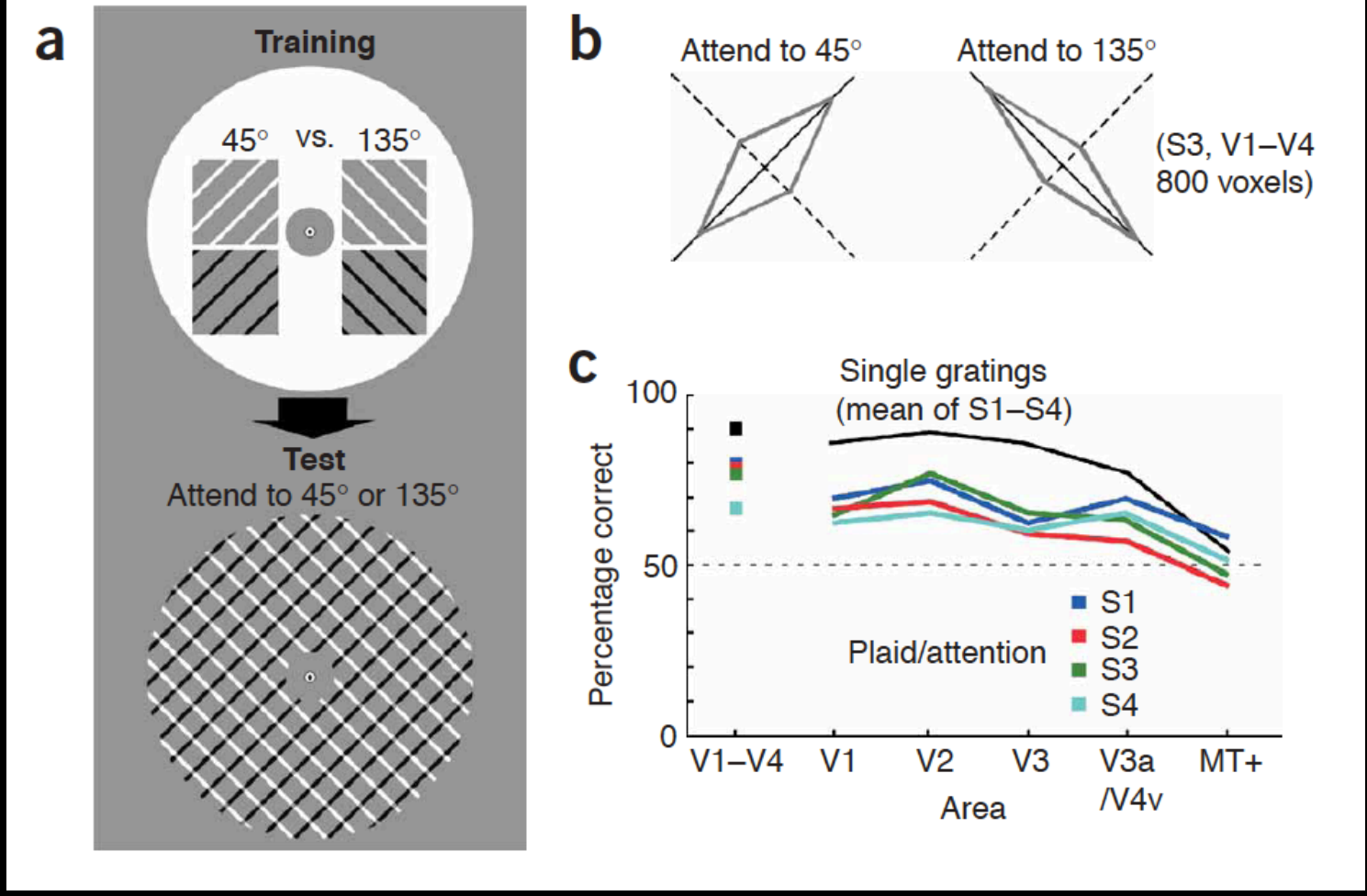
- Participants viewed blocks of oriented lines (8 possible orientations)
- Linear SVM



Highly accurate decoding of orientation



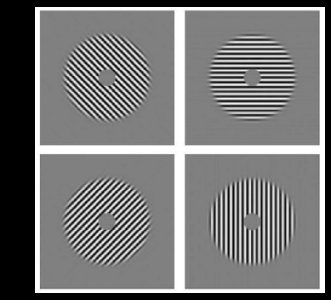
Decoding Attended Orientation



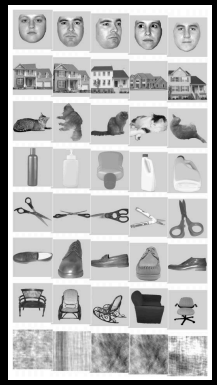
Limitations of Early Decoding Studies

- Restricted stimulus domains
  - Oriented lines
  - Small number of selected categories

Kamitani and Tong (2005)



Haxby et al (2001)

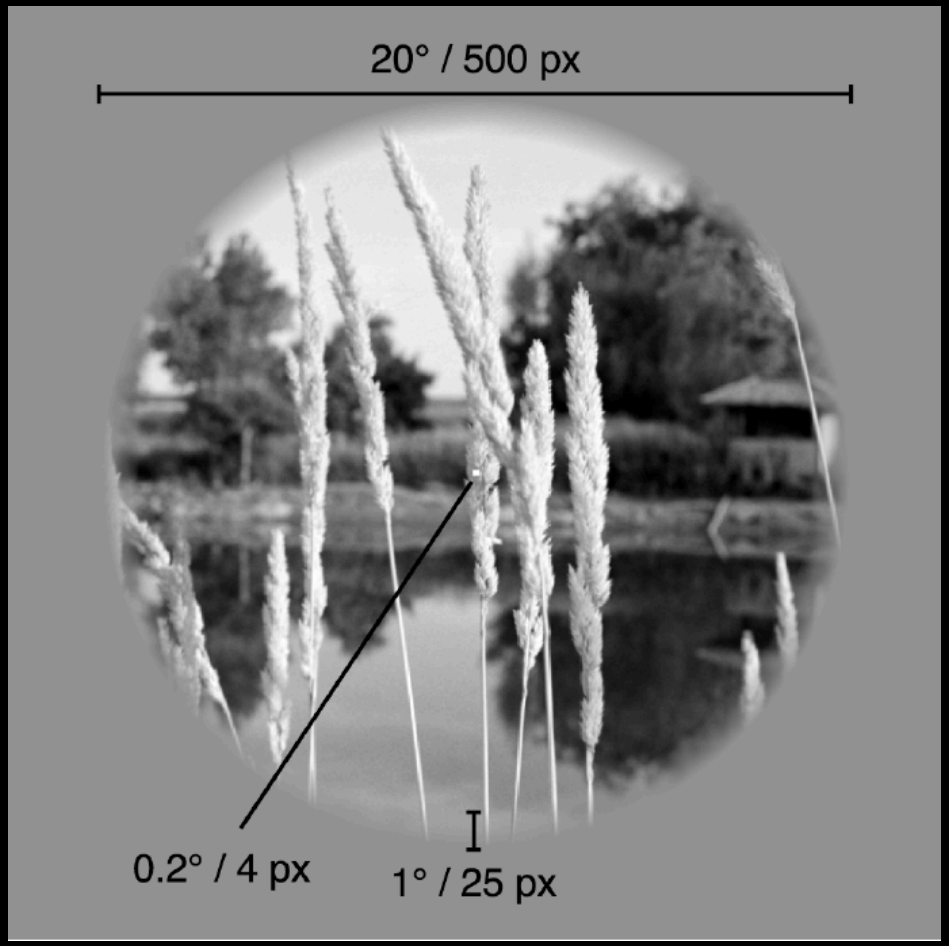


- No decoding of novel stimuli or categories
  [but see Spiridon and Kanwisher(2002)]

Model-based approach to decoding [Kay et al (2008)]

1) Characterize relationship between visual stimuli and fMRI activity (i.e. build a model)
   - Complex, natural visual images
   - Early retinotopic visual cortex
2) Measure fMRI activity to one of many possible novel images
3) Compare actual activity to predicted activity for full set of novel images to determine which image was viewed

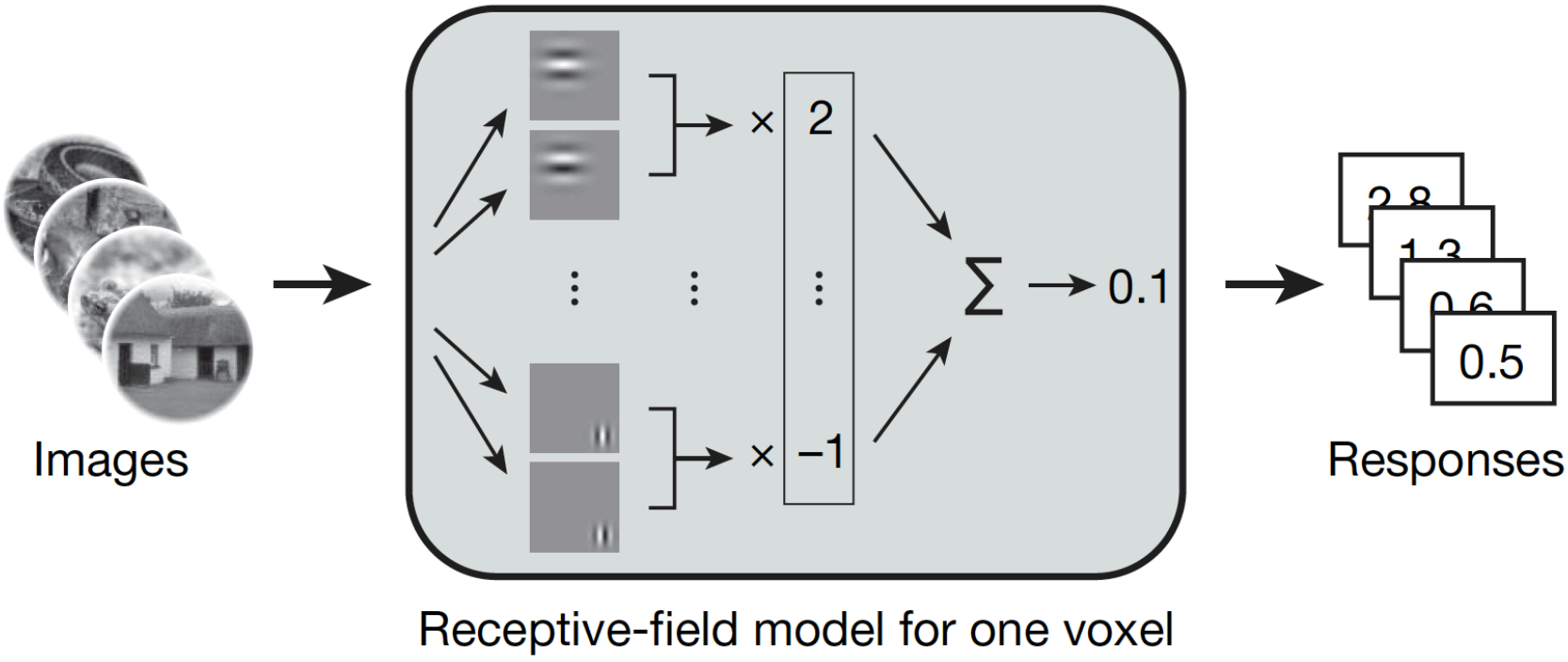
Large gray-scale images



1) Build a Model

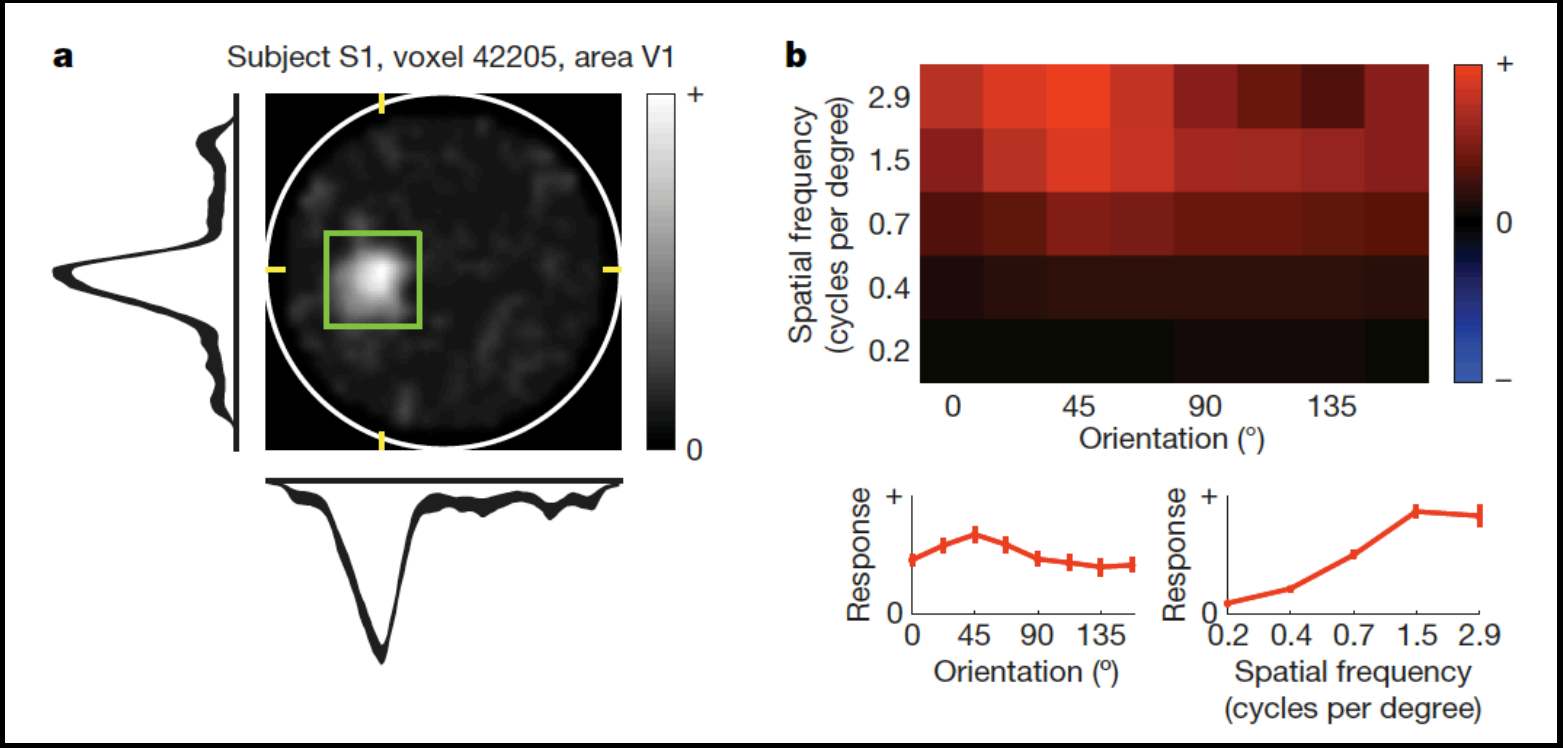
Stage 1: model estimation

Estimate a receptive-field model for each voxel

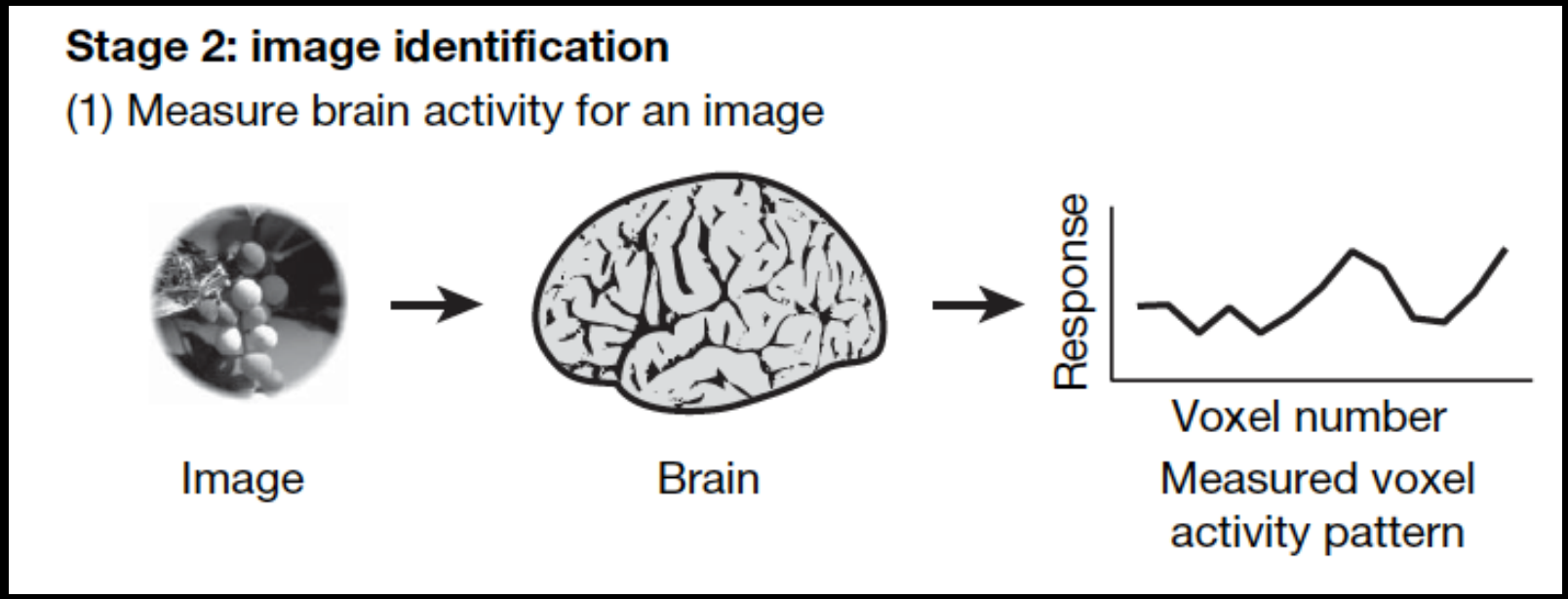


Receptive-field model for one voxel

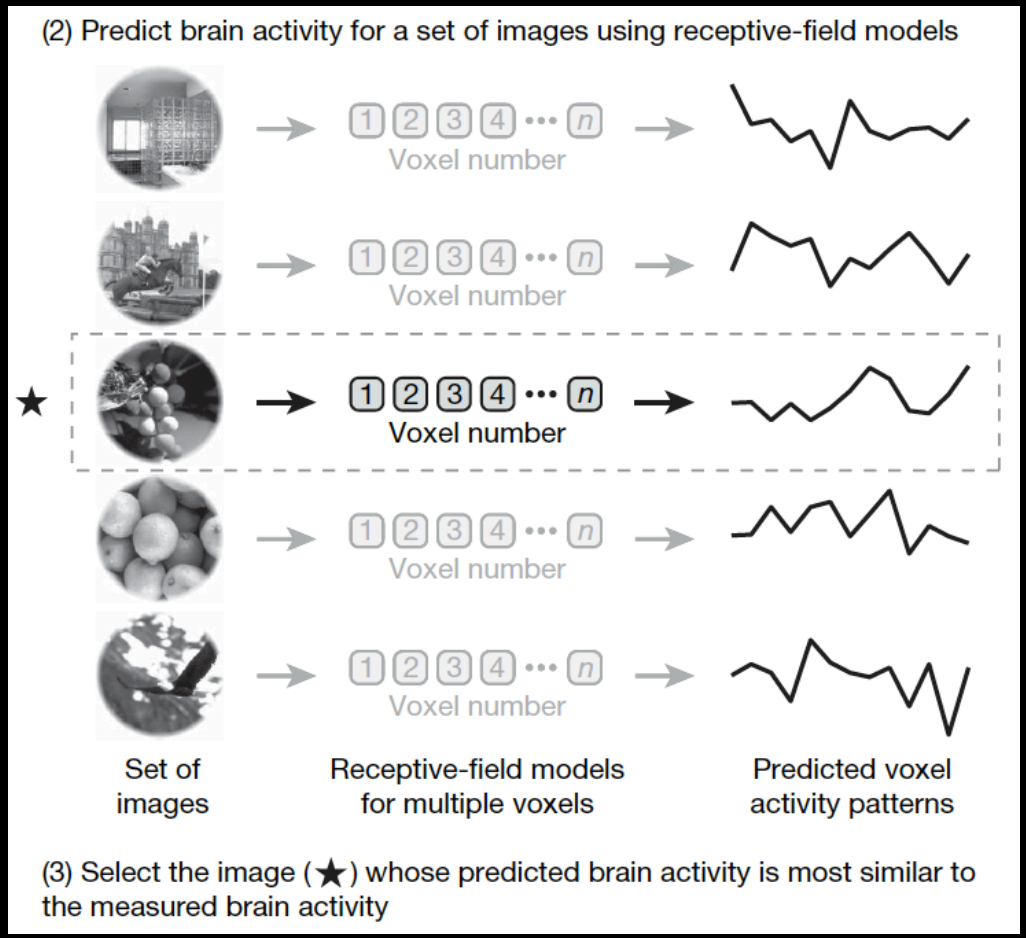
RF model for one voxel



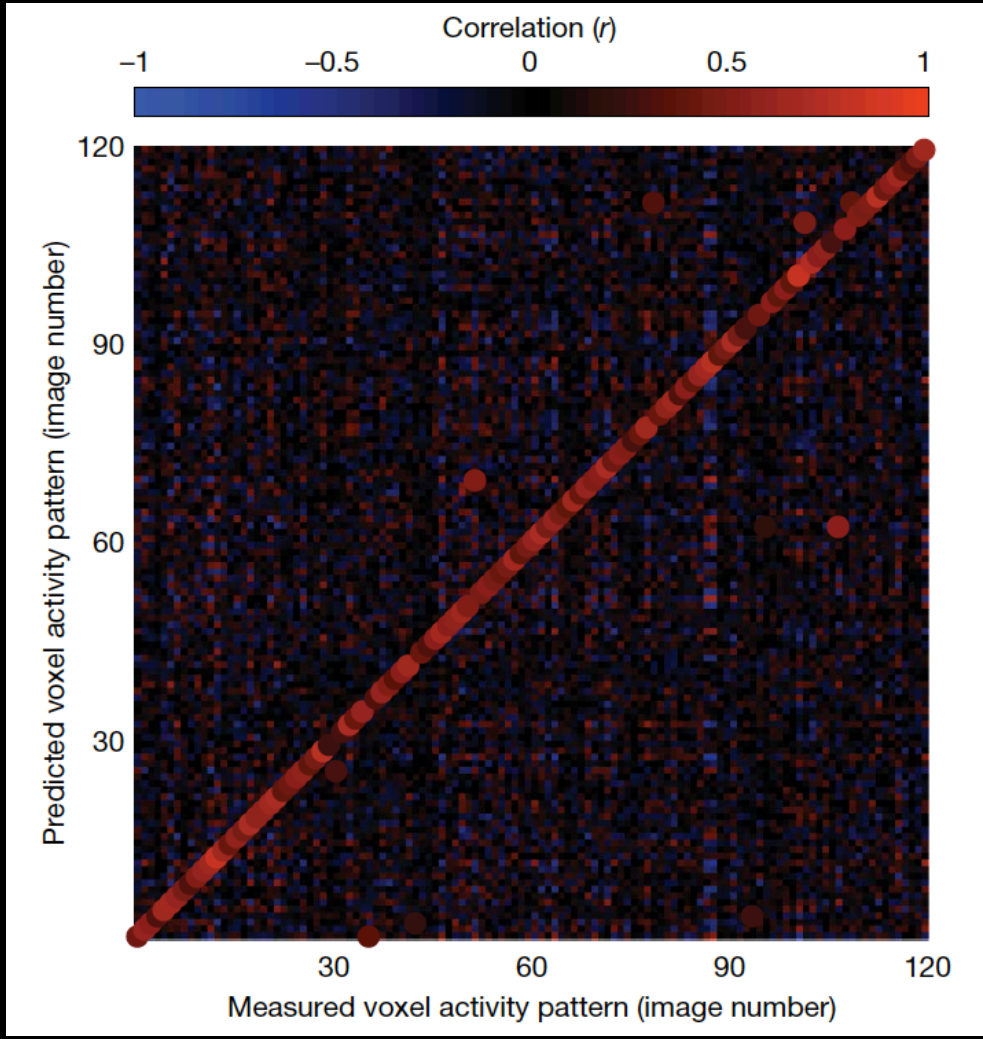
Novel Image to be Identified



Compare observed to predicted activity



Performance



Additional results

- Works on single trials
- Not just retinotopy
- Accurate even with long delay between model fitting and testing

Limitations of Kay et al.

- Still requires comparison with set of candidate images
- Will likely fail with more homogeneous images (e.g. two faces)
- Whole image comparison
  - What about same central object on different backgrounds?
- How sensitive to fixation differences?
- Novel subjects?
- Visual perception is dynamic

Visual Image Reconstruction [Miyawaki et al (2008)]

- Model based decoding
- Characterize relationship between activity and contrast of local image patches
- Use activity to predict contrast within image

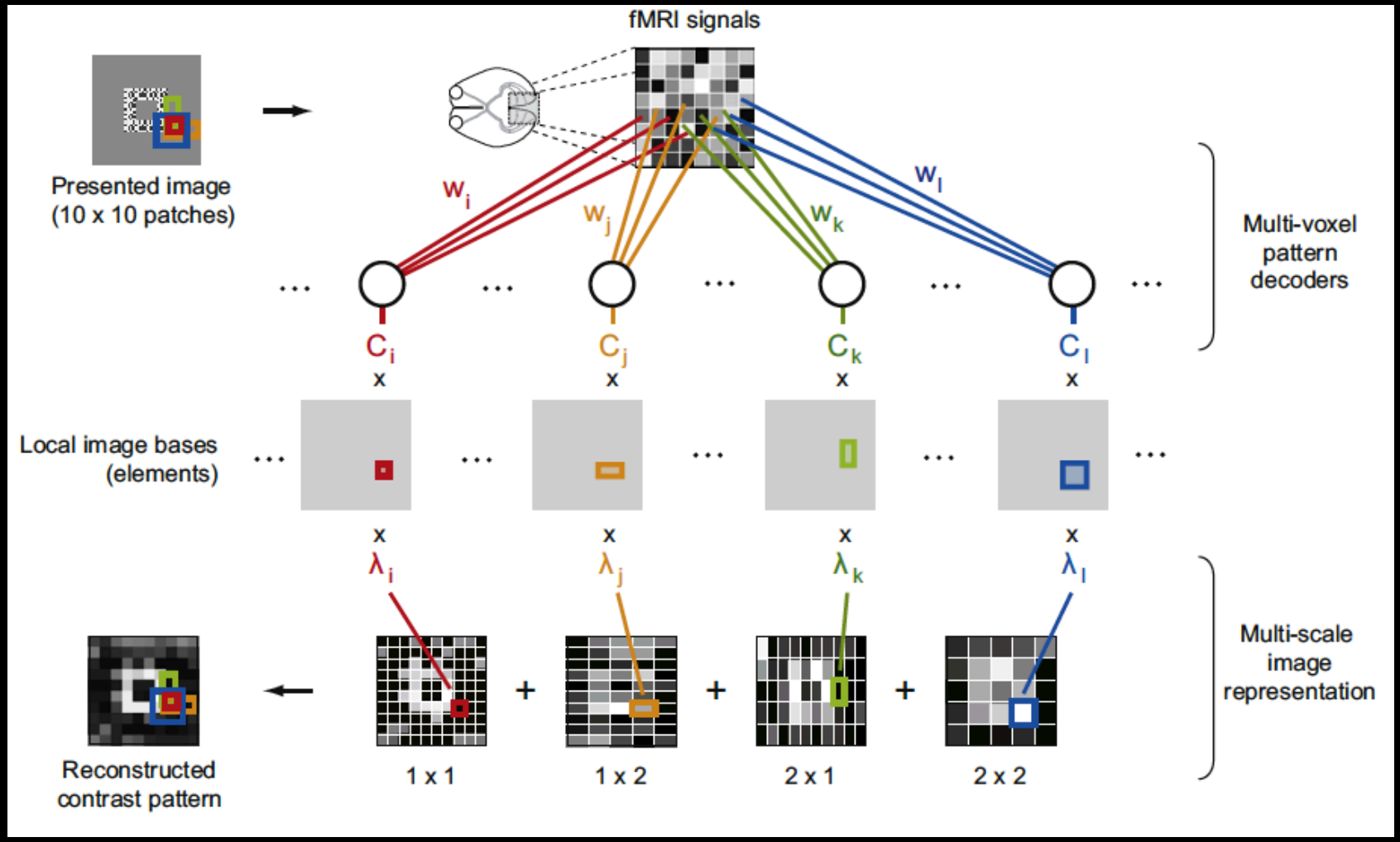
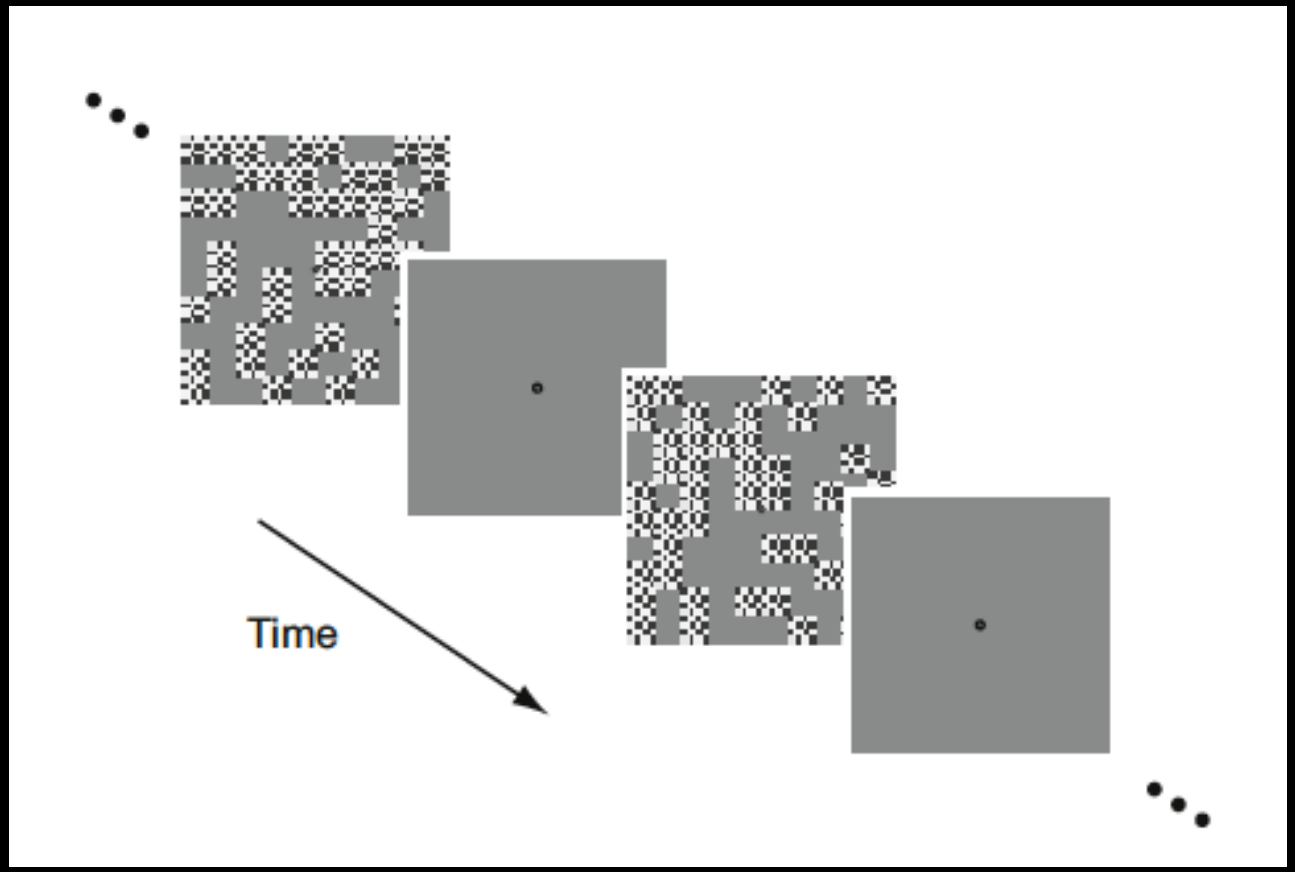
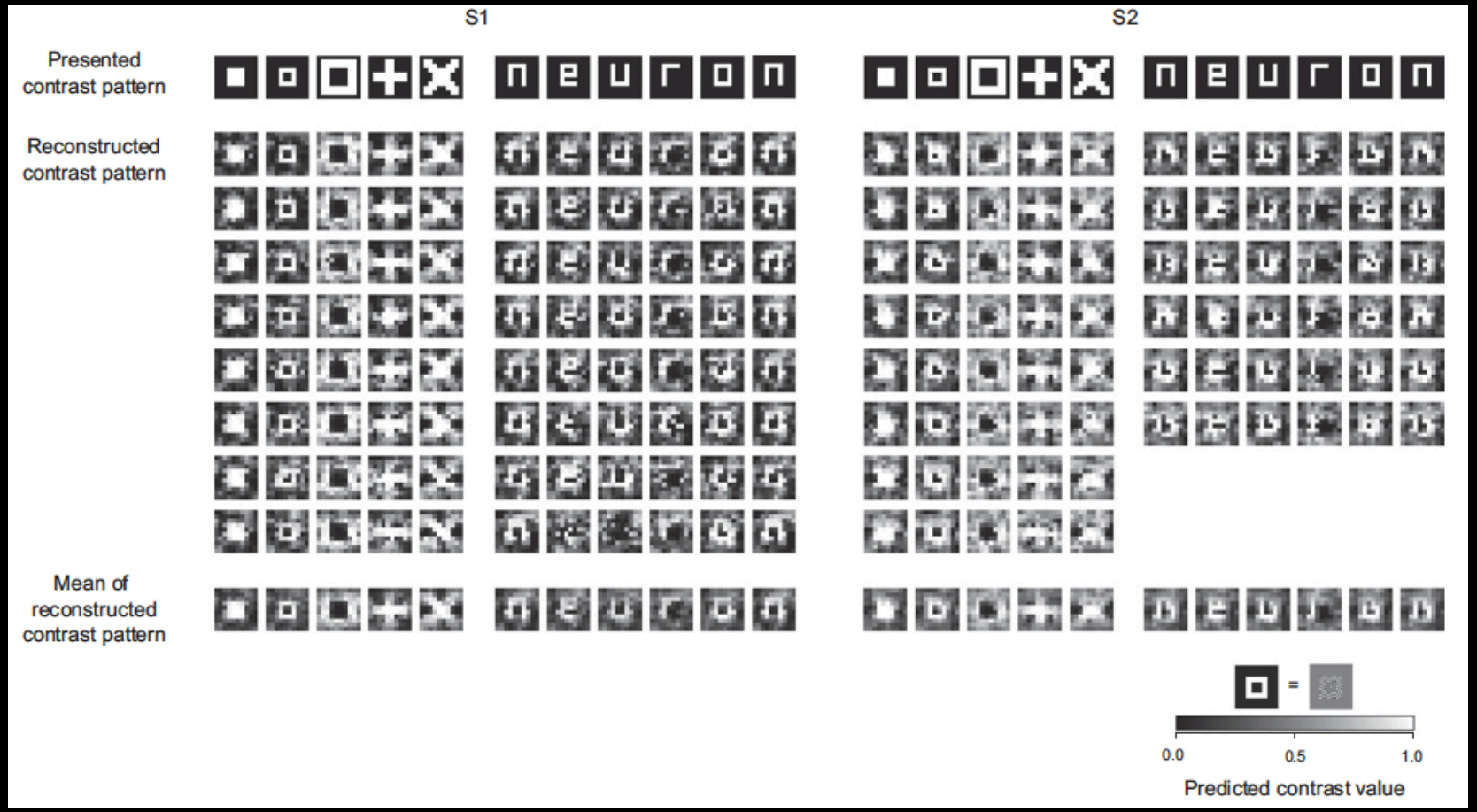

fMRI signals
Presented image
(10 x 10 patches)
w_i
w_j
w_k
w_l
Multi-voxel
pattern
decoders
C_i
C_j
C_k
C_l
x
Local image bases
(elements)
λ_i
λ_j
λ_k
λ_l
Multi-scale
image
representation
Reconstructed
contrast pattern
1 x 1
1 x 2
2 x 1
2 x 2

Image presentation



Reconstructions

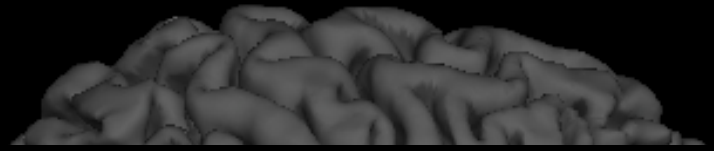


Limitations of Miyawaki et al

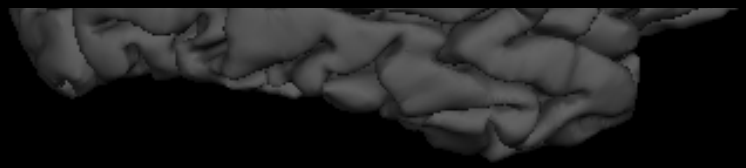
- Similar limitations to Kay et al.
- Simple, non-natural stimuli
- Small image size

For extension of Kay et al. into reconstruction, see Naselaris et al (2009)

“mind reading” “thought identification”



What can we really do?



“prediction” “decoding”

Key Readings

- Cox and Savoy (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19, 261-270.
- Haxby et al. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425-2430.
- Kamitani and Tong (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8, 679-685.
- Kay et al (2008). Identifying natural images from human brain activity. *Nature*, 452, 352-355.
- Kay and Gallant (2009). I can see what you see. *Nature Neuroscience*, 12, 245-246.
- Miyawaki et al (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60, 915-929.
- Mur et al. (2008). Revealing representational content with pattern information fMRI - an introductory guide. *Social Cognitive and Affective Neuroscience*, 4, 101-109.
- Norman et al. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10, 424-430.

Resources

- SVM toolbox
  - http://www.csie.ntu.edu.tw/~cjlin/libsvm/
- Python MVPA toolbox
  - http://www.pymvpa.org/
- Princeton MVPA toolbox
  - http://code.google.com/p/princeton-mvpa-toolbox/